EXHIBIT "E"

Gene, 66 (1988) 45–54
Elsevier

GEN 02398

Nucleotide sequence of the *SUP2* (*SUP35*) gene of *Saccharomyces cerevisiae*

(Omnipotent suppressor; translation ambiguity; gene structure; codon bias analysis; gene homology; elongation factor EF-1α; intron; mitochondrial import)

Vitaliy V. Kushnirov, Michail D. Ter-Avanesyan, Miroslav V. Telckov, Andrei P. Surguchov, Vladimir N. Smirnov and Sergey G. Inge-Vechtomov *

*Institute of Experimental Cardiology, U.S.S.R. Cardiology Research Center, Moscow (U.S.S.R.), and * Department of Genetics, Leningrad State University, Leningrad (U.S.S.R.) Tel. 218-1590*

Received 10 November 1987
Accepted 12 November 1987
Received by publisher 25 February 1988

SUMMARY

A nucleotide sequence of the yeast *Saccharomyces cerevisiae* omnipotent suppressor *SUP2* (*SUP35*) gene is presented. The sequence contains a single open reading frame (ORF) of 2055 bp, which may encode a 76.5-kDa protein. A single transcript of 2.3 kb corresponding to a complete ORF is found. Analysis of codon bias suggests that the *SUP2* gene is not highly expressed. The C-terminal part of the deduced amino acid sequence shows a high homology to yeast elongation factor EF-1α, whereas the N-terminal part is unique for the *SUP2* protein. The N terminus contains a number of short repeating elements and possesses an unusual amino acid composition.

Analysis of the nucleotide and deduced amino acid sequences indicates that three additional proteins could possibly be expressed, two of which might be initiated on internal ATG codons and a third might be formed by alternative splicing. One of these proteins is supposed to be imported into mitochondria. Possible functions of the *SUP2* gene product(s), especially its putative activity as a soluble factor controlling the fidelity of translation, are discussed.

INTRODUCTION

Studies of informational suppression have proved to be useful in elucidating the mechanisms of control of translational fidelity. In all cases studied so far informational suppression results from mutational alterations in the components of protein synthesis apparatus — usually either tRNAs, or ribosomal constituents (Ozeki et al., 1980; Sherman, 1982; Dequard-Chablat et al., 1986; Steege and Söll,

Correspondence to: Dr. V.V. Kushnirov, Institute of Experimental Cardiology, U.S.S.R. Cardiology Research Center, 3rd Cherepkovskaya Street 15a, 121552 Moscow (U.S.S.R.) Tel. 149-0559.

Abbreviations: aa, amino acid(s); bp, base pair(s); EF, elongation factor; kb, kilobases or 1000 bp; MBN, mung-bean nuclease; MBN buffer, see MATERIALS AND METHODS, section b; nt, nucleotide(s); ORF, open reading frame; Pipes, piperazine-*N,N'*-bis[2-ethanesulfonic acid]; PolIk, Klenow (large) fragment of *E. coli* DNA polymerase I; S1 buffer, see MATERIALS AND METHODS, section b; tRNA, transfer RNA; u, unit(s).

1979). Recently, nonsense-suppressor mutations in *tuf* genes coding for EF-Tu have been described in *Escherichia coli* (Vijgenboom et al., 1985). However, similar mutations in eukaryotes have not yet been reported.

For the past several years we were studying recessive omnipotent suppressors in yeast. It was shown that mutations in the genes named *SUP1* (*SUP45*) and *SUP2* (*SUP35*) give rise to a variety of pleiotropic effects, including temperature sensitivity, drug sensitivity and respiratory deficiency. From these observations we concluded that the suppressor genes are essential for viability. Biochemical analysis indicates that suppressor mutations decrease the accuracy of translation and affect protein synthesis both in the cytoplasm and in mitochondria (Surguchov et al., 1984).

Recently, both the *SUP1* and *SUP2* genes were cloned (Breining et al., 1984; Telckov et al., 1986) and the nucleotide sequence of the *SUP1* gene was determined (Breining and Piepersberg, 1986). In this paper we report the nucleotide sequence of the *SUP2* gene. Part of the sequence shows significant homology to yeast EF-1α, suggesting that the *SUP2* gene product is not a canonical ribosomal protein, but rather a soluble translation factor. This essential protein appears to be present in minor quantities and probably has not been detected by biochemical methods. Further characterization of its role may reveal new essential features of the eukaryotic translation machinery.

MATERIALS AND METHODS

(a) Subcloning and sequencing

A shuttle plasmid pSTR4 containing *SUP2* gene (Telckov et al., 1986), was used for the sequence determination. A set of subclones of *SUP2* gene sufficient for sequencing was obtained in two steps: (*i*) restriction fragments of *SUP2* were cloned into M13mp phages (M13mp10, 11, 18, 19), and (*ii*) in some cases subclones were further deleted using DNase I, as described (Lin et al., 1984). DNA restriction, ligation and other enzymatic treatments were carried out according to the suppliers' specifications (Pharmacia P-L Biochemicals). Transformation of *E. coli* (strain JM101) by M13 phage and purification of recombinant phage were done according to Messing (1983). The nucleotide sequence was determined using the dideoxy method of Sanger et al. (1977).

(b) Yeast RNA analysis

Preparation of total yeast RNA was performed, as described by Cottrelle et al. (1985). For the Northern analysis, 20 μg of total RNA were glyoxylated, electrophoresed on an agarose gel and transferred to nitrocellulose (Maniatis et al., 1982). RNA blots were hybridized with strand-specific M13 probes, which were prepared according to Messing (1983).

To obtain a single-stranded ^{32}P-labelled probe for 5′-end mapping, an M13 clone containing fragment *Kpn*I-*Bcn*I (bp 164 to −205, see Fig. 2) with the *Kpn*I site proximal to a sequencing primer site was used. The probe was synthesized by the primer extension with PolIk, cleaved at the 3′ end with *Bcn*I and separated from the template using a 5% polyacrylamide gel (Leer et al., 1984). Hybridization was performed for 6 h at 46°C in 80% formamide, 0.4 M NaCl, 0.4 M Pipes (pH 6.5) and 1 mM EDTA. Total yeast RNA (50 μg) and 100000 cpm of probe were used for each experiment. Then hybridization mixtures were diluted ten-fold with S1 or MBN buffer (30 mM Na · acetate, pH 4.6, 1 mM $ZnSO_4$, 250 mM (50 mM for MBN buffer) NaCl, 20 μg/ml of sonicated and denatured calf thymus DNA), supplemented with 1000–4000 u/ml of the appropriate nuclease and digested for 30 min at 37°C. After chloroform extraction the protected DNA fragments were precipitated with isopropanol in the presence of carrier tRNA and analyzed on a 5% polyacrylamide, 7 M urea sequencing gel.

RESULTS AND DISCUSSION

(a) Nucleotide sequence

For sequence analysis, a shuttle plasmid pSTR4 (Telckov et al., 1986) carrying a minimal fragment of the cloned yeast genomic DNA complementing temperature-sensitive *sup2* mutation was used. The restriction map for this fragment and sequencing

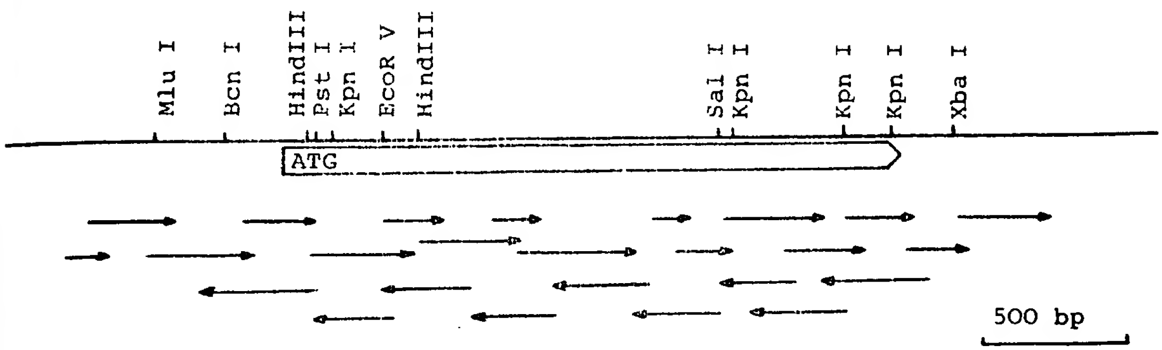

Fig. 1. Restriction map and sequencing strategy for the *SUP2* gene. Position and orientation of the ORF is represented by the arrowed open bar. Small arrows indicate the direction and extent of sequence determination on individual clones.

strategy are shown in Fig. 1. A contiguous sequence of 3320 nt containing a single long ORF of 2055 nt capable of coding for a protein of 76545 Da was determined (Fig. 2).

(b) mRNA analysis

Hybridization of total yeast RNA with a single-stranded M13 probe containing a *Pst*I-*Xba*I fragment revealed a single band of 2.3 kb. The opposite strand of the same fragment did not hybridize with RNA. By S1 and MBN mapping two major transcription start points were found at nt positions −15 and −37, as well as two minor sites at nt positions −57 and −43 (Fig. 3), before the first ATG codon in the ORF. Taking into consideration that the size of the transcript is 2.3 kb, we conclude that the transcript contains the full length of the ORF. The first ATG in the ORF is present on the transcript and therefore it is the most probable initiator of trans-

TABLE I

Codon usage in the *SUP2* gene

aa	Codon[a]		aa	Codon		aa	Codon		aa	Codon	
Phe	TTT	9	Ser	<u>TCT</u>	12	Tyr	TAT	10	Cys	<u>TGT</u>	4
Phe	<u>TTC</u>	7	Ser	<u>TCC</u>	7	Tyr	<u>TAC</u>	25	Cys	TGC	1
Leu	TTA	7	Ser	TCA	6	*ter*[b]	TAA	1	*ter*	TGA	0
Leu	<u>TTG</u>	15	Ser	TCG	3	*ter*	TAG	0	Trp	TGG	4
Leu	CTT	3	Pro	CCT	10	His	CAT	7	Arg	CGT	6
Leu	CTC	0	Pro	CCC	2	His	<u>CAC</u>	6	Arg	CGC	0
Leu	CTA	7	Pro	<u>CCA</u>	18	Gln	<u>CAA</u>	40	Arg	CGA	0
Leu	CTG	3	Pro	CCG	0	Gln	CAG	13	Arg	CGG	0
Ile	<u>ATT</u>	17	Thr	<u>ACT</u>	14	Asn	AAT	24	Ser	AGT	5
Ile	<u>ATC</u>	12	Thr	<u>ACC</u>	16	Asn	<u>AAC</u>	21	Ser	AGC	2
Ile	ATA	3	Thr	ACA	8	Lys	AAA	28	Arg	<u>AGA</u>	11
Met	ATG	19	Thr	ACG	1	Lys	<u>AAG</u>	38	Arg	AGG	1
Val	<u>GTT</u>	26	Ala	<u>GCT</u>	20	Asp	GAT	21	Gly	<u>GGT</u>	45
Val	<u>GTC</u>	10	Ala	<u>GCC</u>	16	Asp	GAC	9	Gly	GGC	10
Val	GTA	9	Ala	GCA	7	Glu	<u>GAA</u>	44	Gly	GGA	3
Val	GTG	5	Ala	GCG	0	Glu	GAG	13	Gly	GGG	2

[a] Codons, preferred in highly expressed yeast genes, are underlined.

[b] Symbol *ter* represents translational stop codons.

```
-738                                                    AGAAATTAAAGCTACTTA
-720 CAACAACGGTCTACTACAAATTAAGGTGCCTAAAATTGTCAATGACACTGAAAAGCCGAAGCCAAAAAAGAGGATCGCCATTGAGGAAATACCCGACGAAGAATTGGAGTTTGAAGAAAA
-600 TCCCAACCCTACGGTAGAAAATTGAATATCGTATCTGTTTATACACACATACATACATTTATATTTATAATAAGCGTTAAAATTTCGGCAGAATATCTGTCAACCACACAAAAATCATAC
                                                                                          Mlu I
-480 AACGAATGGTATATGCTTCATTTCTTTGTTTCGCATTAGCTGCGCTATTTGACTCAAATTATTATTTTTTACTAAGACGACGCGTCACAGTGTTCGAGTCTGTGTCATTTCTTTTGTAAT
-360 TCTCTTAAACCACTTCATAAAGTTGTGAAGTTCATAGCAAAATTCTTCCGCAAAAAGATGAATCTTAGTTCTCAGCCCACCAAAAGAGGTACATGCTAAGATCATACAGAAGTTATTGTC
                                          Bcn I
-240 ACTTCTTACCTTGCTCTTAAATGTACATTACAACCGGGTATTATATCTTACATCATCGTATAATATGATCTTTCTTTATGGAGAAAATTTTTTTTCACTCGACCAAAGCTCCCATTGCT
-120 TCTGAAGAGTGTAGTGTATATTGGTACATCTTCTCTTGAAAGACTCCATTGTACTGTAACAAAAAGCGGTTTCTTCATCGACTTGCTCGGAATAACATCTATATCTGCCCACTAGCAACA

  1  ATG TCG GAT TCA AAC CAA GGC AAC AAT CAG CAA AAC TAC CAG CAA TAC AGC CAG AAC GGT AAC CAA CAA CAA GGT AAC AAC AGA TAC CAA
  1  Met Ser Asp Ser Asn Gln Gly Asn Asn Gln Gln Asn Tyr Gln Gln Tyr Ser Gln Asn Gly Asn Gln Gln Gln Gly Asn Asn Arg Tyr Gln
            Hind III                    Pst I                                          Kpn I
 91  GGT TAT CAA GCT TAC AAT GCT CAA GCC CAA CCT GCA GGT GGG TAC TAC CAA AAT TAC CAA GGT TAT TCT GGG TAC CAA CAA GGT GGC TAT
 31  Gly Tyr Gln Ala Tyr Asn Ala Gln Ala Gln Pro Ala Gly Gly Tyr Tyr Gln Asn Tyr Gln Gly Tyr Ser Gly Tyr Gln Gln Gly Gly Tyr

181  CAA CAG TAC AAT CCC GAC GCC GGT TAC CAG CAA CAG TAT AAT CCT CAA GGA GGC TAT CAA CAG TAC AAT CCT CAA GGC GGT TAT CAG CAG
 61  Gln Gln Tyr Asn Pro Asp Ala Gly Tyr Gln Gln Gln Tyr Asn Pro Gln Gly Gly Tyr Gln Gln Tyr Asn Pro Gln Gly Gly Tyr Gln Gln
                                                                                   EcoR V
271  CAA TTC AAT CCA CAA GGT GGC CGT GGA AAT TAC AAA AAC TTC AAC TAC AAT AAC AAT TTG CAA GGA TAT CAA GCT GGT TTC CAA CCA CAG
 91  Gln Phe Asn Pro Gln Gly Gly Arg Gly Asn Tyr Lys Asn Phe Asn Tyr Asn Asn Asn Leu Gln Gly Tyr Gln Ala Gly Phe Gln Pro Gln
                                                                                            Hind III
361  TCT CAA GGT ATG TCT TTG AAC GAC TTT CAA AAG CAA CAA AAG CAG GCC GCT CCC AAA CCA AAG AAG ACT TTG AAG CTT GTC TCC AGT TCC
121  Ser Gln Gly Met Ser Leu Asn Asp Phe Gln Lys Gln Gln Lys Gln Ala Ala Pro Lys Pro Lys Lys Thr Leu Lys Leu Val Ser Ser Ser

451  GGT ATC AAG TTG GCC AAT GCT ACC AAG AAG GTT GGC ACA AAA CCT GCC GAA TCT GAT AAG AAA GAG GAA GAG AAG TCT GCT GAA ACC AAA
151  Gly Ile Lys Leu Ala Asn Ala Thr Lys Lys Val Gly Thr Lys Pro Ala Glu Ser Asp Lys Lys Glu Glu Glu Lys Ser Ala Glu Thr Lys

541  GAA CCA ACT AAA GAG CCA ACA AAG GTC GAA GAA CCA GTT AAA AAG GAG GAG AAA CCA GTC CAG ACT GAA GAA AAG ACG GAG GAA AAA TCG
181  Glu Pro Thr Lys Glu Pro Thr Lys Val Glu Glu Pro Val Lys Lys Glu Glu Lys Pro Val Gln Thr Glu Glu Lys Thr Glu Glu Lys Ser

631  GAA CTT CCA AAG GTA GAA GAC CTT AAA ATC TCT GAA TCA ACA CAT AAT ACC AAC AAT GCC AAT GTT ACC AGT GCT GAT GCC TTG ATC AAG
211  Glu Leu Pro Lys Val Glu Asp Leu Lys Ile Ser Glu Ser Thr His Asn Thr Asn Asn Ala Asn Val Thr Ser Ala Asp Ala Leu Ile Lys

721  GAA CAG GAA GAA GAA GTG GAT GAC GAA GTT GTT AAC GAT ATG TTT GGT GGT AAA GAT CAC GTT TCT TTA ATT TTC ATG GGT CAT GTT GAT
241  Glu Gln Glu Glu Glu Val Asp Asp Glu Val Val Asn Asp Met Phe Gly Gly Lys Asp His Val Ser Leu Ile Phe Met Gly His Val Asp

811  GCC GGT AAA TCT ACT ATG GGT GGT AAT CTA CTA TAC TTG ACT GGC TCT GTG GAT AAG AGA ACT ATT GAG AAA TAT GAA AGA GAA GCC AAG
271  Ala Gly Lys Ser Thr Met Gly Gly Asn Leu Leu Tyr Leu Thr Gly Ser Val Asp Lys Arg Thr Ile Glu Lys Tyr Glu Arg Glu Ala Lys

901  GAT GCA GGC AGA CAA GGT TGG TAC TTG TCA TGG GTC ATG GAT ACC AAC AAA GAA GAA AGA AAT GAT GGT AAG ACT ATC GAA GTT GGT AAG
301  Asp Ala Gly Arg Gln Gly Trp Tyr Leu Ser Trp Val Met Asp Thr Asn Lys Glu Glu Arg Asn Asp Gly Lys Thr Ile Glu Val Gly Lys
```

lation. It is interesting to note that sequences surrounding the two major transcription start points are similar to each other (Fig. 4).

(c) 5′- and 3′-flanking regions

A promoter element TATATT is located in a position typical for such elements in yeast, i.e., between bp −105 and −98 before the first ATG. The sequence AATAAA, which is thought to be a eukaryotic polyadenylation signal (Fitzgerald and Shenk, 1981), is situated 84–89 bp downstream from the terminating TAA. The sequence TAG...TAGT...TTT, a potential transcription termination signal in yeast (Zaret and Sherman, 1982), was found 115–142 bp downstream from the termination codon TAA. An interesting feature of the 3′-flanking region is the presence of the repeats $(TA)_{11}$ (95 bp downstream from TAA) and $(CAT)_{11}$ (350 bp downstream from TAA).

(d) Codon usage

The *SUP2* gene differs markedly in codon usage from highly expressed yeast genes, particularly ribosomal protein genes (Table I). The codon bias index according to Bennetzen and Hall (1982) was determined to be 0.42, whereas the range of values for ribosomal proteins is 0.79–0.94 (Sharp et al., 1986). Such a difference could mean, according to estimates given by Bennetzen and Hall (1982), that *SUP2* mRNA is at least an order of magnitude less abundant than mRNAs of ribosomal protein genes.

(e) Deduced amino acid sequence

A part of the amino acid sequence beginning with Met-254 (the third methionine in the sequence) is homologous to the full length of yeast EF-1α (Kushnirov et al., 1987). The remaining N-terminal part can be divided near the second methionine into

```
 991 GCC TAC TTT GAA ACT GAA AAA AGG CGT TAT ACC ATA TTG GAT GCT CCT GGT CAT AAA ATG TAC GTT TCC GAG ATG ATC GGT GGT GCT TCT
 331 Ala Tyr Phe Glu Thr Glu Lys Arg Arg Tyr Thr Ile Leu Asp Ala Pro Gly His Lys Met Tyr Val Ser Glu Met Ile Gly Gly Ala Ser

1081 CAA GCT GAT GTT GGT GTT TTG GTC ATT TCC GCC AGA AAG GGT GAG TAC GAA ACC GGT TTT GAG AGA GGT GGT CAA ACT CGT GAA CAC GCC
 361 Gln Ala Asp Val Gly Val Leu Val Ile Ser Ala Arg Lys Gly Glu Tyr Glu Thr Gly Phe Glu Arg Gly Gly Gln Thr Arg Glu His Ala

1171 CTA TTG GCC AAG ACC CAA GGT GTT AAT AAG ATG GTT GTC GTC GTA AAT AAG ATG GAT GAC CCA ACC GTT AAC TGG TCT AAG GAA CGT TAC
 391 Leu Leu Ala Lys Thr Gln Gly Val Asn Lys Met Val Val Val Val Asn Lys Met Asp Asp Pro Thr Val Asn Trp Ser Lys Glu Arg Tyr

1261 GAC CAA TGT GTG AGT AAT GTC AGC AAT TTC TTG AGA GCA ATT GGT TAC AAC ATT AAG ACA GAC GTT GTA TTT ATG CCA GTA TCC GGC TAC
 421 Asp Gln Cys Val Ser Asn Val Ser Asn Phe Leu Arg Ala Ile Gly Tyr Asn Ile Lys Thr Asp Val Val Phe Met Pro Val Ser Gly Tyr

1351 AGT GGT GCA AAT TTG AAA GAT CAC GTA GAT CCA AAA GAA TGC CCA TGG TAC ACC GGC CCA ACT CTG TTA GAA TAT CTG GAT ACA ATG AAC
 451 Ser Gly Ala Asn Leu Lys Asp His Val Asp Pro Lys Glu Cys Pro Trp Tyr Thr Gly Pro Thr Leu Leu Glu Tyr Leu Asp Thr Met Asn
         SalI                                                           Kpn I
1441 CAC GTC GAC CGT CAC ATC AAT GCT CCA TTC ATG TTG CCT ATT GCC GCT AAG ATG AAG GAT CTA GGT ACC ATC GTT GAA GGT AAA ATT GAA
 481 His Val Asp Arg His Ile Asn Ala Pro Phe Met Leu Pro Ile Ala Ala Lys Met Lys Asp Leu Gly Thr Ile Val Glu Gly Lys Ile Glu

1531 TCC GGT CAT ATC AAA AAG GGT CAA TCC ACC CTA CTG ATG CCT AAC AAA ACC GCT GTG GAA ATT CAA AAT ATT TAC AAC GAA ACT GAA AAT
 511 Ser Gly His Ile Lys Lys Gly Gln Ser Thr Leu Leu Met Pro Asn Lys Thr Ala Val Glu Ile Gln Asn Ile Tyr Asn Glu Thr Glu Asn

1621 GAA GTT GAT ATG GCT ATG TGT GGT GAC CAA GTT AAA CTA AGA ATC AAA GGT GTT GAA GAA GAA GAC ATT TCA CCA GGT TTT GTA CTA ACA
 541 Glu Val Asp Met Ala Met Cys Gly Glu Gln Val Lys Leu Arg Ile Lys Gly Val Glu Glu Glu Asp Ile Ser Pro Gly Phe Val Leu Thr

1711 TCG CCA AAG AAC CCT ATC AAG AGT GTT ACC AAG TTT GTA GCT CAA ATT GCT ATT GTA GAA TTA AAA TCT ATC ATA GCA GCC GGT TTT TCA
 571 Ser Pro Lys Asn Pro Ile Lys Ser Val Thr Lys Phe Val Ala Gln Ile Ala Ile Val Glu Leu Lys Ser Ile Ile Ala Ala Gly Phe Ser
                                                                                Kpn I
1801 TGT GTT ATG CAT GTT CAT ACA GCA ATT GAA GAG GTA CAT ATT GTT AAG TTA TTG CAC AAA TTA GAA AAG GGT ACC AAC CGT AAG TCA AAG
 601 Cys Val Met His Val His Thr Ala Ile Glu Glu Val His Ile Val Lys Leu Leu His Lys Leu Glu Lys Gly Thr Asn Arg Lys Ser Lys

1890 AAA CCA CCT GCT TTT GCT AAG AAG GGT ATG AAG GTC ATC GCT GTT TTA GAA ACT GAA GCT CCA GTT TGT GTG GAA ACT TAC CAA GAT TAC
 631 Lys Pro Pro Ala Phe Ala Lys Lys Gly Met Lys Val Ile Ala Val Leu Glu Thr Glu Ala Pro Val Cys Val Glu Thr Tyr Gln Asp Tyr
                                             Kpn I
1980 CCT CAA TTA GGT AGA TTC ACT TTG AGA GAT CAA GGT ACC ACA ATA GCA ATT GGT AAA ATT GTT AAA ATT GCC GAG TAA ATTTCTTGCAAACA
 661 Pro Gln Leu Gly Arg Phe Thr Leu Arg Asp Gln Gly Thr Thr Ile Ala Ile Gly Lys Ile Val Lys Ile Ala Glu ...

2073 AAGTAAATGCAAACACAATAATACCGATCATAAAGCATTTTCTTCTATATTAAAAACAAGGTTTAATAAAGCTGTTATATATATATATATATATATAGACGTATAATTAGTTTAGTTC
                                             Xba I
2193 TTTTGTACCATATACCATAAACAAGGTAAACTTCACCTCTCAATATATCTAGAATTTCATAAAAATATCTAGCAAGGTTTCAACTCCTTCAATCACGTTTTCATCATAACCCTTCCCCG

2313 CGTTATTTCAGAATGTGCAAAATCTATTAGTGACATGGAACTCAAAGAACCAGTTGTTTTTTTGTCCTTTGGTCCTTCGCTGCTTCCCTCGGCATCATCATCATCATCATCATCATTAT

2433 ATCATCGTCGTCATCATCGTCTATAAAATCATCTCGCATAAGTTTGTCAACATCATTTAGTAATTCCCATCGCTCCGGGTCTCCTTCGTAAATAAACAAAAGACTACTTGATATCATTC

2553 AACTTCTTCTTCTAGCATAGTATTATAAAA
```

Fig. 2. Nucleotide sequence and deduced amino acid sequence of the *SUP2* gene. The location of restriction sites is indicated. Sequence elements TATATT, AATAAA and TAG...TAGT...TTT (Zaret and Sherman, 1982), which may be relevant for initiation or termination of the transcription, are underlined by solid lines. Major and minor transcription start points are marked by downward arrows. Underlined by dashed lines are: HOMOL1-like sequence, the second and third in-frame ATG codons and sequences GTATGT and TACTAAC typical for yeast introns. Second ATG and the GTATGT sequence do overlap.

two fragments, both having an unusual amino acid composition (Table II; Fig. 5).

Region A is a region of 123 aa, beginning at the first methionine and contains repeats of three sequence elements, which make up most of its length (Fig. 6). Sequence Gln-Gly-Gly-Tyr-Gln-(Gln)-Gln-Tyr-Asn-Pro is repeated about six times (Fig. 6b). This region is rich in Gln (28%), Gly (17%), Asn (16%) and Tyr (16%), all four amino acids making up 78%.

Region B is a region of aa 124–253 rich in charged amino acids, Lys (18%) and Glu (18%), which may be further subdivided into four stretches: (*1*) a stretch of aa 124–164 is positively charged and resembles the signal sequences for mitochondrial import (von Heijne, 1976; see RESULTS, section g, for details); (*2*) a stretch of aa 165–222 contains several repeats of tetrapeptides: Lys-Lys-Glu-Glu, Thr-Lys-Glu-Pro, Glu-Glu-Lys-Ser, Thr-Glu-Glu-Lys; (*3*) a stretch of aa 223–235 does not contain charged residues; (*4*) a stretch of aa 236–253 carries a negative charge (9 aa residues out of 18 are Asp or Glu).

It is interesting that region B contains 24 Lys, but does not contain Arg.

(f) Possible existence of additional *SUP2* gene products

RNA analysis reveals a single transcript for the *SUP2* gene containing a complete ORF. However, detailed analysis of the nucleotide and deduced amino acid sequences points to the possibility of existence of shorter transcripts and corresponding

protein products. One may suggest a possibility of translation initiation on the second and third ATG codons as well as excision of the part of the coding sequence resulting from splicing, as it is shown in sections **g–i**, below.

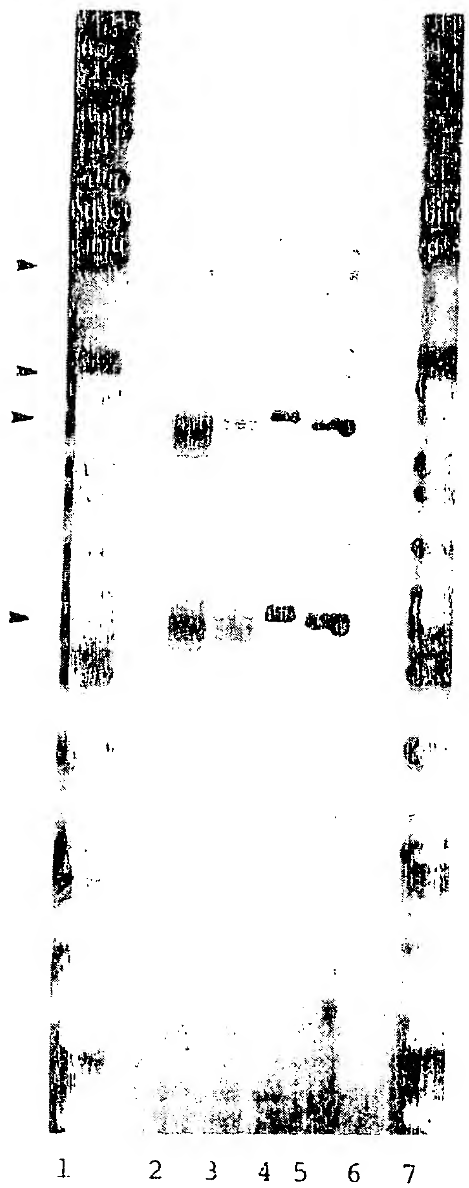

Fig. 3. Mappi[illegible] he 5′ end of the *SUP2* mRNA. T[illegible] east RNA was hybri[illegible]ed to a single-stranded ^{32}P-labeled *Kpn*I [illegible] fragment (nt 1[illegible] o −205) at 46°C, in 80% formamide and treated with S[illegible] nuclease in the following concent[illegible]as: lanes: 2, 2000 [illegible] S1; 3, 4000 u/ml of S1; 4, 1000 [illegible] of MBN; 5, 4000 [illegible] of MBN, 6, control without yeast RN[illegible] 000 u/ml of S1. A [illegible] y sequencing lane of a known sequ[illegible] was used as a mar[illegible] (lanes 1 and 7). The transcription st[illegible] ints are indicated [illegible] owheads.

(g) Initiation of the translation on the second in-frame ATG

As we have shown earlier, many *sup2* mutations cause a respiratory deficiency, reduction in mitochondrially synthesized cytochrome content and decrease in the rate of protein synthesis in mitochondria. These data allowed to predict the existence

```
              -37↓
              TCGACTTGCTCGGAA
Consensus:    T  AYYTGCYCR  A
              TATATCTGCCCACTA
              -15↑
```

Fig. 4. Similarity of the two major transcription start point regions. Transcription start points are marked by arrows. R designates purine (A or G), Y designates pyrimidine (T or C). The upstream sequence is on top, the downstream sequence is at the bottom (see Fig. 2), and the consensus sequence is in the middle line.

TABLE II

Amino acid composition of the *SUP2* gene[a]

aa	Region			The entire *SUP2* protein
	A 1–123	B 124–253	E 253–685	1–685
Ala	6	9	28	43
Arg	2	0	16	18
Asn	20	7	18	45
Asp	2	7	21	30
Cys	0	0	5	5
Gln	35	6	12	53
Glu	0	23	34	57
Gly	21	2	37	60
His	0	1	12	13
Ile	0	3	29	32
Leu	1	7	27	35
Lys	1	24	41	66
Met	1	1	17	19
Phe	3	1	12	16
Pro	6	8	16	30
Ser	5	10	20	35
Thr	0	11	28	39
Trp	0	0	4	4
Tyr	20	0	15	35
Val	0	10	40	50
Total	123	130	432	685

[a] Amino acid composition of regions A (aa 1–123), B (aa 124–253), E (aa 254–685) and the entire *SUP2* protein is shown. Unusually high content of some amino acids is underlined.

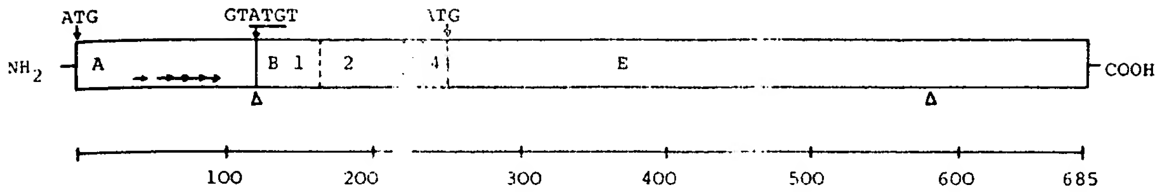

Fig. 5. Schematic representation of the predicted primary structure of the *SUP2* protein. Segment (A) represents a region containing extensive repeats of 8–10-bp sequence elements (shown by short horizontal arrows), and rich in Gln (28%), Gly (17%), Asn (16%) and Tyr (16%). Segment (B) represents a region with high content of charged residues Lys (18%) and Glu (18%). It may be subdivided into four stretches: (1) a positively charged stretch similar to signal sequences for mitochondrial import (von Heijne, 1986); (2) a stretch containing several tetrapeptide repeats rich in Lys and Glu; (3) a stretch without charged residues; (4) a stretch carrying a negative charge. Segment (E) represents a region homologous to the full length of yeast EF-1α. The sites corresponding to the first, second and third ATG codons are denoted by downward arrows. Putative splice sites are marked by triangles. The scale below the map is in aa. (See also Fig. 2 and Table II.)

of a *SUP2* gene product, which may be imported into mitochondria (Surguchov et al., 1984). Most of such imported proteins have a signal sequence at their N termini, which is positively charged and able to form an amphiphilic helix (von Heijne, 1986). A similar sequence element is present in a single site in the *SUP2* protein after the second methionine (aa 124–164).

Conserved 12-bp sequences, HOMOL1 and RPG, are present in the 5′-flanking regions of most of yeast ribosomal protein genes (Teem et al., 1984; Leer et al., 1985). The sequence HOMOL1 is also found in 5′-flanking regions of genes encoding EF-1α (Huet et al., 1985) and in the *SUP1* gene (Breining and Piepersberg, 1986). It has been proposed, that these sequences are required for the transcriptional regulation of the components of the translational apparatus (Huet et al., 1985). The HOMOL1 box is usually located before the TATA box at a distance of 150–400 bp upstream from the transcription start point. In the *SUP2* gene a sequence AACATC-TATATC similar to the HOMOL1 sequence, AACATC(T/C)(G/A)T(A/G)CA, is present. However, since this sequence is situated after presumed TATA box at nt positions −27 to −16 before the first ATG codon, it is possible that it regulates initiation of transcription at a site before the second ATG. A corresponding putative TATA box is located 150 bp upstream from the second ATG.

(h) Initiation of translation on the third in-frame ATG

Upon alignment of homologous regions of amino acid sequences of EF-1α and the *SUP2* protein-initiating methionine of the EF-1α corresponds exactly to the third methionine in the *SUP2* protein, thus indicating possible involvement of the latter in the initiation of translation. The following observation confirms this suggestion. Upon deletion from the *SUP2* gene of a restriction fragment *Hin*dIII-*Hin*dIII (nt 99–434), the second in-frame ATG is removed and the reading frame beginning from the first ATG is disrupted. However, high copy number plasmids containing such deletions still complement certain temperature-sensitive *sup2* mutations. This result can be explained only by the existence of a protein initiated on the third in-frame ATG. The calculated M_r of this protein is 48039.

A minor mRNA band of 1.4 kb, hybridizing to the coding segment of the *SUP2* gene, which has been observed by Surguchov et al. (1986), may correspond to a transcript initiated before the third ATG codon. However, such a band was not found in this study. A possible explanation for this discrepancy is that this transcript occurs in relatively small amounts depending on the conditions, used for growth.

(i) Possibility of alternative splicing

The contiguous ORF of the *SUP2* gene contains sequences that are typical for introns in *S. cerevisiae* genes, including a completely conserved sequence TACTAAC, which is present in all yeast introns. This sequence is found in the *SUP2* gene around bp 1700. At the 5′ end of yeast introns a sequence GTATGT or less frequently, GTACGT is located. Both of these sequences are found in the ORF of the *SUP2* gene near bp 364 and 1046, respectively. A trinucleotide TAG (bp 1748) nearest to the sequence

A

```
M S D S N Q G N N Q Q N Y Q Q Y S Q N G N Q Q Q G N N R Y Q    30
G Y Q A Y N A Q A Q P A G G Y Y Q N Y Q G Y S G Y Q Q G G Y    60
Q Q Y N P D A G Y Q Q Q Y N P Q G G Y Q Q Y N P Q G G Y Q Q    90
Q F N P Q G G R G N Y K N F N Y N N N L Q G Y Q A G F Q P Q   120
S Q G M S L N D F Q K Q Q K Q A A P K P K K T L K L V S S S   150
G I K L A N A T K K V G T K P A E S D K K E E E K S A E T K   180
E P T K E P T K V E E P V K K E E K P V Q T E E K T E E K S   210
E L P K V E D L K I S E S T H N T N N A N V T S A D A L I K   240
E Q E E E V D D E V V N D M   254
```

B

```
                     Q P    41
 42   a G G Y y Q N Y Q G   51
 57   Q G G Y Q - Q Y N P   65
 66   D A G Y Q Q Q Y N P   75
 76   Q G G Y Q - Q Y N P   84
 85   Q G G Y Q Q Q F N P   94
 95   Q G G r g - N Y      101
112       G Y Q a g F Q P  119
120   Q
```

6. Analysis of the amino acid sequence of the *SUP2* gene. (A) Deduced structure of the N-terminal region of the *SUP2* protein. one-letter amino acid notation is used. Repeat sequences are overlined and numbered. Charged residues are marked with (+) or . (B) Alignment of the most extensive repeat element. One-aa gaps (dashes) were introduced in some places. Conservative amino ls are given in capitals.

CTAAC may be regarded as a 3′ end of this pothetical intron. The first of two donor splice es (GTATGT) seems to be a more likely candidate the 5′ end of the intron, because in this case the ding frame is not shifted by splicing. Location of s site is not random. The GTATGT sequence covers the second ATG codon and the border of the A and B regions of the deduced polypeptide (Fig. 5). The size of the protein product corresponding to spliced mRNA would be 25 kDa.

A large and functionally important part of the sequence lies inside the proposed intron, including

an assumed signal for mitochondrial import and domains, for which participation in GTP- and aminoacyl-tRNA-binding is predicted due to homology with EF-1α (Kushnirov et al., 1987). This indicates that the unspliced transcript must be expressed, even if splicing does occur. Such alternative splicing was not described in the yeast *S. cerevisiae*.

Although we did not detect multiple transcripts of the *SUP2* gene, their existence as well as expression of the corresponding protein products cannot be excluded.

(j) Possible functions of the *SUP2* protein

The C-terminal part of the *SUP2* protein, beginning from the third methionine (Met-254), shows significant homology to yeast EF-1α as well as to a family of analogous factors from other species (Kushnirov et al., 1987). The degree of amino acid homology amounts to 62%, considering conservative amino acid substitutions as homologous. Furthermore, nonhomologous stretches of significant length are absent in this region. This allows us to suggest that the *SUP2* protein possesses, apart from its N-terminal domains, the same functional domains as EF-1α, including GTP- and aminoacyl-tRNA-binding domains, where the degree of homology is highest. One might speculate then that these two proteins act at the same site on the ribosome and that their mode of action is rather similar. However, it is important to emphasize, that they are not interchangeable, since disruption of the *SUP2* gene is lethal (M.D.T.-A. and A.R. Dagkesamanskaya, in preparation). Furthermore, as pointed out in RESULTS, section d, the *SUP2* protein appears to be much less abundant, than EF-1α. Taken together, these and other data are consistent with the assumption that the product of the *SUP2* gene is a soluble factor that participates in the control of the fidelity of translation.

In a well-studied translation system of *E. coli*, a minor protein similar to the *SUP2* protein has not been found yet. At the same time, omnipotent suppressor mutations in EF-Tu have been described (Vijgenboom et al., 1985). Moreover, analysis of these mutants revealed a reduction in the accuracy of the protein synthesis at both the primary aminoacyl-tRNA selection and the proofreading steps (Tapio and Kurland, 1986). This allows us to suggest that EF-Tu, apart from a function analogous to EF-1α, may also perform a proofreading function, for which the *SUP2* protein is specialized in *S. cerevisiae*.

To determine the role of *SUP2* gene product(s) in protein synthesis it will be necessary to identify and purify the protein and study it biochemically. Genetic approaches and recombinant DNA techniques may also give valuable information, for example, examination of nucleotide substitutions, leading to suppression.

ACKNOWLEDGEMENTS

We are grateful to Dr. G.P. Samokhin and Dr. A.B. Sudarickov for the help with computer analysis of the data.

REFERENCES

Bennetzen, J.L. and Hall, B.D.: Codon selection in yeast. J. Biol. Chem. 257 (1982) 3026–3031.

Breining, P. and Piepersberg, W.: Yeast omnipotent suppressor *SUP1* (*SUP45*): nucleotide sequence of the wild-type and a mutant gene. Nucl. Acids Res. 14 (1986) 5187–5197.

Breining, P., Surguchov, A.P. and Piepersberg, W.: Cloning and identification of a DNA fragment coding for the *SUP1* gene of *Saccharomyces cerevisiae*. Curr. Genet. 8 (1984) 467–470

Cottrelle, P., Thiele, D., Price, V.L., Memet, S., Micouin, J.Y., Marck, C., Buchler, J.M., Sentenac, A. and Fromageot, P.: Cloning, nucleotide sequence and expression of one of two genes, coding for yeast elongation factor 1α. J. Biol. Chem. 260 (1985) 3090–3096.

Dequard-Chablat, M., Coppin-Raynal, E., Picard-Bennoun, M and Madjar, J.J.: At least seven ribosomal proteins are involved in the control of translational accuracy in a eukaryotic organism. J. Mol. Biol. 190 (1986) 167–175.

Fitzgerald, M. and Shenk, T.: The sequence 5′–AAUAAA–3 forms part of the recognition site for polyadenylation of late SV-40 messenger RNAs. Cell 24 (1981) 251–260.

Huet, J., Cottrelle, P., Cool, M., Vignais, M.-L., Thiele, D. Marck, C., Buchler, J.M., Sentenac, A. and Fromageot, P. A general upstream binding factor for genes of the yeast translational apparatus. EMBO J. 4 (1985) 3539–3547.

Kushnirov, V.V., Ter-Avanesyan, M.D., Surguchov, A.P. Smirnov, V.N. and Inge-Vechtomov, S.G.: Localisation of possible functional domains in *SUP2* gene product of the yeast *Saccharomyces cerevisiae*. FEBS Lett. 215 (1987 257–260.

Leer, R.G., Van Raamsdonk-Duin, M.M.C., Mager, W.H. and Planta, R.G.: The primary structure of the gene encoding yeast ribosomal protein L 16. FEBS Lett. 175 (1984) 371–376.

Leer, R.G., Van Raamsdonk-Duin, M.M.C., Mager, W.H. and Planta, R.G.: Conserved sequences upstream of yeast ribosomal protein genes. Curr. Genet. 9 (1985) 273–277.

Lin, H.C., Lei, S.P. and Wilcox, G.: An improved sequencing strategy. Anal. Biochem. 147 (1985) 114–119.

Maniatis, T., Fritsch, E.F. and Sambrook, J.: Molecular Cloning. A Laboratory Manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, 1982, pp. 200–201.

Messing, J.: New M13 vectors for cloning. Methods Enzymol. 101 (1983) 20–78.

Ozeki, H., Inokuchi, H., Yamao, F., Kodaira M., Sakano, H., Ikemura, T. and Shimura, Y.: Genetics of nonsense suppressor transfer RNAs in *Escherichia coli*. In Söll, D., Abelson, J.N., Schimmel, P.R. (Eds.), Transfer RNA: Biological Aspects. Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, 1980, pp. 341–362.

Sanger, F., Nicklen, S. and Coulson, A.R.: DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA 74 (1977) 5463–5467.

Sharp, P.M., Tuohy, T.M.F. and Mosurski, K.R.: Codon usage in yeast: cluster analysis clearly differentiates highly and lowly expressed genes. Nucl. Acids Res. 14 (1986) 5125–5143.

Sherman, F.: Suppression in the yeast *Saccharomyces cerevisiae*. In Strathern, J.N., Jones, E.W. and Broach, J.R. (Eds.), Molecular Biology of the Yeast *Saccharomyces*: Metabolism and Gene Expression. Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, 1982, pp. 463–486.

Steege, D.A. and Söll, D.G.: Suppression. In Goldberger, R.F. (Ed.), Biological Regulation and Development, Vol. I. Plenum, New York, 1979, pp. 433–485.

Surguchov, A.P., Smirnov, V.N., Ter-Avanesyan, M.D. and Inge-Vechtomov, S.G.: Ribosomal suppression in eukaryotes. In Skulachev, V.P. (Ed.), Physicochemical Biology Reviews, Vol. 4. Soviet Scientific Reviews, Moscow, 1984, pp. 147–205.

Surguchov, A.P., Telckov, M.V. and Smirnov, V.N.: Absence of structural homology between *SUP1* and *SUP2* genes of yeast *Saccharomyces cerevisiae* and identification of their transcripts. FEBS Lett. 206 (1986) 147–150.

Tapio, S. and Kurland, C.G.: Mutant EF-Tu increases missense error in vitro. Mol. Gen. Genet. 205 (1986) 186–188.

Teem, J.L., Abovich, N., Kaufer, N.F., Schwindinger, W.F., Warner, J.R., Levy, A., Woolford, J., Leer, R.J., Van Raamsdonk-Duin, M.M.C., Mager, W.H., Planta, R.J., Schultz, L., Friesen, J.D., Fried, H. and Roshbash, M.: A comparison of yeast ribosomal protein gene DNA sequences. Nucl. Acids Res. 12 (1984) 8295–8312.

Telckov, M.V., Surguchov, A.P., Dagckesamanskaya, A.R. and Ter-Avanesyan, M.D.: Isolation of a chromosomal DNA fragment containing *SUP2* gene of the yeast *Saccharomyces cerevisiae*. Genetika 22 (1986) 17–25 (in Russian).

Vijgenboom, E., Vink, T., Kraal, B. and Bosch, L.: Mutants of elongation factor EF-Tu, a new class of nonsense suppressors. EMBO J. 4 (1985) 1049–1052.

von Heijne, G.: Mitochondrial targeting sequences may form amphiphilic helices. EMBO J. 5 (1986) 1335–1342.

Zaret, K.S. and Sherman, F.: DNA sequence required for efficient transcription termination in yeast. Cell 28 (1982) 563–573.

Communicated by A.A. Bayev.